Bandolier 37

What do we think?
What do we know?
What can we prove?

Evidence-based health care **£2.50** **March 1997 Volume 4 Issue 3**

Trust me, I'm a .....

If you ride a motorbike from London to Glasgow, your risk of dying is about 300 times greater than if you fly or take a coach. ***Bandolier*** has found some more detailed figures that many of you asked for (page 8). That is a **voluntary** risk you are taking. You put your trust in the pilot or coach driver. Much of medicine involves trust, and a theme of ***Bandolier's*** stories this month is the risk, the peril, of misplaced trust.

Put not your trust in princes

If you have a mole excised you trust a pathologist to make an accurate judgement as to whether it is benign or malignant. That judgement may be more difficult than we punters realise (page 2). Recognising that there is uncertainty at every stage of the diagnostic process is hard for professionals. It's good to see that recognition in print with a will to improve.

Trusting reviews

Systematic reviews do not always tell exactly the same story, or necessarily the whole truth. Examples of four reviews of TENS (page 3) make interesting reading because of the nuances in different methods and conclusions. Another example is the way that NNT calculations can extract more useful truth from studies of new anti-epileptic drugs [British Medical Journal 1997 314: 603].

Urge to purge

Decisions not to use prostate-specific-antigen (PSA) for prostate cancer screening (page 7) are based on the argument that why bother to screen if there is no effective diagnosis or treatment? That is a statement which those diagnosed with prostate cancer, and those looking after them, may not like to hear. The clash between the politics of screening and shop floor treatment resonates again. PSA can be used successfully as a marker in treatment, so a knee jerk response to the reports of 'purging' all PSA tests would be inappropriate.

Delete as inappropriate

Going to the chemist with a prescription ***Bandolier*** was surprised to be given a form to fill in. The form asked for details of ***Bandolier's*** ailments and medications. Ostensibly there is a laudable safety motive here, a double-check that patients do not take medicines which clash or are inappropriate. This information could also be misused, as for marketing purposes, or confidential details could leak. Are there any safeguards? ***Bandolier*** did not fill in the form.

Pathology as art appreciation

Have you ever considered just how much a pathologist is like an art expert? Just as a Lord Clarke (of Civilisation) could spot a 17th century Italian masterpiece, know a Cezanne because of the particular character of the blues in it, or pick out an original Seurat from a pile of copies because of the nature of the little blobs the artist employed, so a pathologist has to recognise one set of visual impressions as being cancer or not.

Histopathology has long been regarded as providing a "gold standard" diagnosis against which all others have to be measured. This gold standard is falling in value after being tested and found to be less than 24 carat. There are two main weaknesses - both pretty fundamental - inter-observer variability and the lack of statistically-based predictive power of their "diagnoses".

Impressionism in pathology

Inter-observer variability (how different pathologists interpret a slide) has to be distinguished from intra-observer variability (how the same pathologist interprets the same slide at different times). Even if the same specimen is shown to the same observer there is not always perfect agreement between the two opinions.

The major problem though is the variability between observers. When any paper assesses the accuracy of a diagnostic test which depends upon an individual's perception, on their vision or hearing, or on some subjective decision-making process, it is essential to ensure that the study includes an assessment of the degree of inter-observer variability. Any such paper should be using a statistical technique, of which the KAPPA is the most widely used, to assess the degree of the inter-observer variability.

In this issue

Agreement by chance

Techniques such as kappa allow for agreement by chance, taking into account the prevalence of the condition being assessed, so the degree of agreement in those areas where agreement or disagreement matters can be assessed. Kappa (or κ) has values between 0 (a random effect) and 1 (perfect agreement). In practice a kappa score of greater than 0.4 is taken to indicate that agreement is becoming reasonable, while 0.6 or above is good agreement.

A recent editorial in the Journal of Pathology [1] pointed out that apart from grouping lesions into broad categories (benign versus malignant), poor agreement is not infrequent. Some examples from the literature (albeit generally with small numbers of pathologists, but with perhaps more attention given than is the norm in clinical practice) quoted in the editorial are reproduced in the table below.

Diagnosis of melanoma

As part of the National Institutes of Health consensus conferences on the diagnosis and treatment of early melanoma, a study was conducted to review pathology specimens and measure inter-pathologist agreement [2]. A panel of eight pathologists expert in melanoma diagnosis was selected. Each submitted five cases (slide plus clinical history and eventual diagnosis), which they considered "classic" cases of melanomas or melanocytic nevi that shared histological features with melanomas.

From these 37 cases were selected and anonymised. Slides with histories (but not eventual diagnosis) were then randomised independently of the organising group. The slides were sent to each panel member in turn, so that the same glass slides were used by each panel member. Each case was to be scored as benign, malignant, or indeterminate. Any description other than this had to be defined.

Results

All 37 cases were reviewed, without loss or breakage. Eight benign cases and five malignant cases were agreed unanimously. Lack of unanimous agreement occurred in 24 cases (62%). Two or more discordant diagnoses were made in 14 cases (38%) and discordance was three or more in 8 cases (22%). The kappa was 0.5, indicating only moderate agreement.

It is illuminating to look at the extremes. One expert (and these were all experts in melanoma, don't forget) thought 21 cases were malignant and 16 were benign. Another thought 10 were malignant, 26 benign, and one indeterminate. Between them, these two pathologists disagreed on 12 out of 37 cases, and in 11 cases one pathologist identified a case as malignant while the other identified the same case as benign.

An accompanying editorial [3] speaks of "shattered illusions". It comments that "the conclusions of the article....should be chilling not only to physicians, but to patients, and sobering to lawyers for plaintiffs."

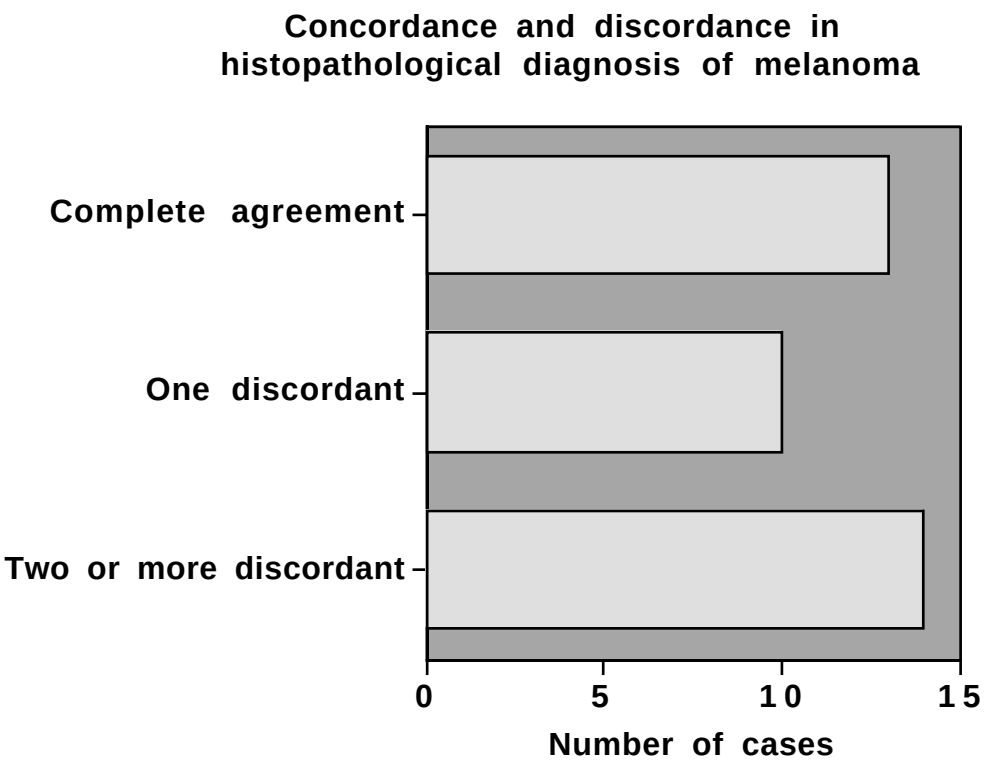
Names or diagnoses

Even if there is perfect agreement that a collection of cells or piece of tissue can be given a certain name with a high level of agreement, the name itself may be meaningless. The name may just not convey that a person is at risk of disease or has a disease. In the past these names were called "diagnoses". A diagnosis implies that a patient has a disease with a certain prognosis, established scientifically. This is not always the case. A name may be just a label given to appearances that histopathologists name that way - like "impressionism" or "cubism" - not a diagnosis like "tuberculosis".

Results of studies on agreement between histopathologists

Organ Feature	Agreement	Kappa
Liver - piecemeal necrosis	47%	0.2
Rectal cancer - grading	50-69%	0.1 - 0.5
Lymph node - Hodgkin's classification	56%	0.4
Cervical intra-epithelial neoplasia	no data	0.01 - 0.5
Grading anal intra-epithelial neoplasm	no data	0.1 - 0.6
Liver transplant acute rejection	no data	0.3 - 0.5
Breast cancer classification	73%	0.4
Breast cancer grade		0.3
Breast cancer invasive sub-type		0 - 0.3
Breast cancer atypical hyperplasia		0.2

References are in Fleming, 1996 [1].



An excellent leader in The Lancet [4] emphasised the danger of labelling something with a name which might have harmful consequences for the patient, both physical and psychological, without any benefit. The title, of which ***Bandolier*** would have been proud, was "Carcinoma-in-situ of the breast: have pathologists run amok?" The author, a pathologist in New Mexico, delivers a trenchant attack on "in-situ-diagnosis" and suggests that pathologists' monopoly on giving names to things they see is challenged. He argues that "their diagnoses of carcinoma-in-situ have transmitted more fear than knowledge into the clinical arena". He calls for "terminology consumers" to start saying what names and classifications they would find most helpful. He finishes up with the striking sentence that "it does not require specialist training in pathology to recognise that the patient's diagnosis should not be an anachronism sustained by anecdotes, conjecture and tradition".

Eels in grease

Diagnostic tests, of whatever type and done in any specialty, have their problems. Tackling these problems (and particularly trying to describe ways in which they can be dealt with) is like trying to hold on to eels in grease. Similar problems occur in other areas, and in diagnosis done in test tubes. ***Bandolier*** was moved to run this article because of the realistic attitude being taken by pathologists about their profession and by the forceful attitudes expressed in both the British and American pathology journals. Do our readers have examples of other diagnostic tests, of whatever sort, that can lead us astray or do extremely well?

References:
1 KA Fleming. Evidence-based pathology. Journal of Pathology 1996 179: 127-8.
2 ER Farmer, R Gonin, MP Hanna. Discordance in the histopathologic diagnosis of melanoma and melanocytic nevi between expert pathologists. Human Pathology 1996 27: 528-31.
3 AB Ackerman. Discordance among expert pathologists in diagnosis of melanocytic neoplasms. Human Pathology 1996 27:1115-6
4 E Foucar. Carcinoma-in-situ of the breast: have pathologists run amok? Lancet 1996 347: 707-8.

DOES TENS WORK?

Having one systematic review on a topic is great. Having two is terrific, but four would seem to be an embarrassment of riches. Or at least it would be an embarrassment if, having come at a topic from different perspectives, they reached conflicting conclusions. For transcutaneous electrical nerve stimulation (TENS) for treating pain, by and large they do not.

Is TENS much used?

According to the Canadian technology assessment paper [1] in which the authors surveyed TENS use across Canada, the answer would seem to be yes, it is widely used. After surveying 50 hospitals with 200 or more beds they estimated over 450,000 uses of TENS take place in Canadian hospitals each year with widespread use in physiotherapy and for acute pain (used by 93% of hospitals), labour and delivery (43%), and for chronic pain (96%).

TENS in acute pain

The Canadian study [1] included randomised and non-randomised studies, though it split them for descriptive analysis. They reviewed 39 studies in postoperative pain, dental pain, dysmenorrhoea and cervical pain. They described results as varied, but you have to read awfully closely to find much good about TENS. The review suggested that the authors of the original papers which were randomised reached positive conclusions about TENS in 19/34 papers, but a negative or balanced conclusion in 15/34.

In a review limited to randomised studies in acute postoperative pain [2], TENS was judged by the reviewers to be no better than placebo in 15 out of 17 randomised studies. Of 19 trials with pain outcomes which were not randomised, 17 of 19 the authors of the original papers had concluded that TENS had a beneficial effect. This is another good example of bias in non-randomised studies (see ***Bandolier*** 17).

	TENS effective	TENS ineffective
Randomised	2	15
Inadequately randomised or not randomised	17	2

TENS in labour pain

The Canadian review [1] summarised 6/9 randomised trials as reaching negative conclusions. This is a similar result to the second review [3], which examined eight reports, of which five were judged to have a negative result with TENS no better than placebo or sham-TENS. However, the three studies which were judged to be positive were positive only on weak outcomes like additional pain relieving measures and increased time to epidural local anaesthetic.

Additional analgesic interventions were significantly less likely with the use of TENS [3], with a number-needed-to-treat of 14 (95%CI 7 - 119) for one woman in labour to be spared an epidural or intramuscular injection. Of the four trials which reported this outcome, only the two smaller trials (23 women receiving TENS) were statistically significant, while the two larger studies (208 women receiving TENS) were not significantly different from placebo.

A randomised trial of 94 women in the first stage of labour published since the reviews were done [4] undermines even this possible level of benefit. It found no difference in analgesic requirement between active TENS and disabled TENS equipment, nor any difference on pain scores.

TENS in chronic pain

The Canadian review found 20 randomised trials, of which nine were definitely positive for TENS on some measure, but eight were negative for TENS.

A disappointing review limited to chronic low back pain [5] included just six papers. Two of the original studies were of electroacupuncture, which is not the same as TENS. One trial had only ten patients randomised, six patients to TENS and four to placebo. TENS was not significantly different from placebo. Electroacupuncture was significantly better than placebo, but in only two studies with 30 patients given electroacupuncture.

What is one to make of all this?

First of all there are methodological considerations. Adequate blinding of TENS is extremely difficult, so that most trials should best be regarded as open even if described a blinded. This must confer some degree of bias towards TENS.

Second, there was a tendency in all these reviews to point to an overall lack of methodological rigour in the original studies (but acknowledging that these trials are difficult). The trials with the best methods tended to produce negative results.

Then there is the issue of statistical validity. Put simply, many of these trials make a number of different measurements, only some of which show statistical benefit. So choosing just those measurements which are significant, and ignoring those that are not significant, can give a spurious weight to a review. This is especially true when statistical benefit is of dubious clinical value. Reviewers and readers should always make up their own minds, not just take a headline result chosen because of statistics.

So our reading of these reviews should be sceptical, especially when, as in acute postoperative pain, there are adequate alternatives. For labour pain there may just be an argument for good quality trials which examine the issue of delay or avoidance of interventions like epidurals or intramuscular opiates which carry some risk to mother or baby.

Babies and bathwater

Chronic pain is a different matter. Where the evidence is not clear cut, where some patients are seen to benefit, and where alternatives may not work for all patients, then carrying on using TENS until there is some clarification makes sense. That does put some heat on getting well-designed studies of sufficient power to provide practically useful answers underway.

In the meantime, those of you who see full-page adverts in the national newspapers full of happy souls extolling the virtues of TENS might like to refer the Advertising Standards Authority to these reviews.

References:

1 J Reeve, D Menon, P Corabian. Transcutaneous electrical nerve stimulation (TENS): a technology assessment. International Journal of Technology Assessment 1996 12:2: 299-324.

2 D Carroll, M Tramèr, H McQuay, B Nye, A Moore. Randomization is important in studies with pain outcomes: systematic review of transcutaneous electrical nerve stimulation in acute postoperative pain. British Journal of Anaesthesia 1996 77: 798-803.

3 D Carroll, M Tramèr, H McQuay, B Nye, A Moore. Transcutaneous electrical nerve stimulation in labour pain: a systematic review. British Journal of Obstetrics & Gynaecology 1997 104: 169-175.

4 JM van der Ploeg, HAM Vervest, AL Liem, JH Schagen van Leeuwen. Transcutaneous nerve stimulation (TENS) during the first stage of labour: a randomized clinical trial. Pain 1996 68: 75-8.

5 JG Gadsby, MW Flowerdew. The effectiveness of transcutaneous electrical nerve stimulation (TENS) and acupuncture-like transcutaneous electrical nerve stimulation (ALTENS) in the treatment of patients with chronic low back pain. Cochrane Library 1997 issue 1.

CALCIUM AND VITAMIN D FOR PREVENTING HIP FRACTURES

One of the topics covered in the ***Bandolier*** conference on osteoporosis was that of supplementing the diets of elderly at-risk populations with calcium and vitamin D. Some health authorities have large populations of elderly nursing home residents who are mobile but fragile, and formulating a policy on this may be important to such an authority.

There is accumulated evidence that positive calcium balance - promoted by adequate calcium and vitamin D intake - is beneficial in the prevention of bone loss in older people. Stronger bones should result, and stronger bones should be more resistant to fracture if older people fall.

The question is whether positive action to ensure adequate dietary calcium results in fewer broken bones. Chapuy and colleagues [1] have undertaken a randomised controlled trial to test that. The Chapuy study has been criticised because it is a selected group, of women only, which might not reflect the whole elderly population (the average age, for instance, was in the mid-80s), but it probably does reflect any similar UK population of elderly but mobile people living in homes.

Randomised trial

The trial took place in France in a selected group of elderly women ranging in age from 69 to 106 years at entry. Randomisation was between 1.2 grams of elemental calcium (as an aqueous suspension of tricalcium phosphate) plus 20 µg vitamin D3 daily at lunch time and matched placebos. Vitamin D3 was included in the trial because dairy produce in France is not fortified with vitamin D.

The inclusion criteria were:-

- Living in nursing home or apartment house for elderly people.
- Ambulatory - walking outdoors easily to walking indoors with stick.
- No serious medical condition.
- Life expectancy of at least 18 months.

Exclusions were:

- Taking corticosteroids, thyroxine or anticonvulsants within last year.
- Treated with fluoride salts for more than three months.
- Treated with vitamin D or calcium in last six months.

Assessments

Clinical status and dietary calcium intake was assessed at baseline. They were then re-assessed at 6, 12 and 18 months when they clinical status was recorded, together any information about fractures or adverse events. A further assessment was made 36 months after the trial started [2].

The main outcome measures were hip fractures and all non-vertebral fractures, though a small group of women also had biochemical variables and bone density measured.

Evaluation

Evaluation was made on three bases:-

- For women treated and followed up for the full 18 (or 36) months.
- For women who had received treatment for varying lengths of time when they had a fracture, dropped out, or died (active-treatment analysis).
- For all women (intention to treat), but censoring data by excluding those who died from causes other than fracture.

Results

There was a significantly reduced number of fractures (all fractures and hip fractures) for all methods of analysis at 18 months and at 36 months. Odds ratios are not given here as they are not particularly informative.

1728 women were treated and followed for the full 18 months. Dropout rates were similar in the two groups. Adverse events causing drop-out were mainly gastrointestinal (nausea, diarrhoea or epigastric pain); they occurred in 68 women with a similar distribution between the two groups.

The number-needed-to-treat to prevent one fracture at 18 months and at 36 months is given in the table on page 6 with 95% confidence intervals. For hip fractures, the NNT of 20 means that 20 women have to be treated with calcium and vitamin D for 3 years to prevent a hip fracture in one who would have had a fractured hip had they all not been treated. The NNT to prevent fractures at any non-vertebral site was 14.

NNT values at 36 months were about twice as good as those at 18 months, indicating that continued treatment increases the benefit. Over longer periods, it is likely that the NNT would continue to fall (see [3]).

The data on fractures is buttressed by information from bone density measurements - an average 3% increase in the treated group compared with a 5% decrease in the untreated group. Biochemical measures (reduced PTH and increased vitamin D) also supported improved bone metabolism.

What are the holes in these studies? Few, actually. The data are there for the most conservative of determinants of efficacy. It might be possible to quibble about the inclusion criteria, because a small minority of elderly women will be treated with corticosteroids, thyroxine or anticonvulsants, but not many. No men were studied, probably because their lifetime risk (5%) is about half that of women (12%).

Controversy

A systematic review of vitamin D supplementation in the Cochrane Library [3] comes to a different conclusion. It calculates the NNT at 36 months as 40 (95% CI 22 - 210). The reason for the difference seems to be that these reviewers have taken an "all patients randomised" view of the denominator for the intention to treat analysis, whereas the in original pa-

per information from women who died from causes other than fracture were censored. Increasing the denominator without increasing the numerator will give larger NNTs and less good effects.

This is an interesting issue. There is a purist line which would support not censoring, but in a population where all cause death rates are high, reducing hip fractures in those still living is clearly the clinical outcome desired. Perhaps the key lesson is that if *different* interventions were to be compared to prevent hip fracture, then they should be compared on the same basis, whether that was censored or not censored data. This may be particularly important when it comes to cost-effectiveness analysis.

Cost effective?

Depends, especially with the NNT consideration above.

The cost of a broken hip has been estimated at £5,000 (strictly hospital medical costs [4]), but others have suggested much higher figures. At the Bandolier conference, figures as high as £17,500 were suggested. Prophylactic treatment of high-risk patients would be worthwhile if the cost of prophylaxis was about the same or less than the cost of treating a fractured hip. The combination used in the Chapuy study was Ostram, costing about 20p a day or £73 per person for a year.

With a NNT of 20, twenty patients would have to be treated for three years to prevent a hip fracture in one of them, so the cost of prevention would be £4,400.

With a NNT of 40, but accepting that some people would not use three year's worth of supplement, the cost would be likely to be less than £8,800 - perhaps £6,600 in an elderly population.

So it looks as if there is a balance between spending money on supplements or spending money on treating broken hips.

But that is just the cost basis. Most elderly people with broken hips have a much reduced quality of life, over and above the trauma and hurt. They rarely get back to the same degree of independent living as they enjoyed before they fell and broke their hip. Where there is an identifiable population at risk, purchasers may want to do their own computations for benefit in their own area.

Who is most at risk?

Of patients admitted with hip fracture, 87% are over the age of 65, and 82% of these are women.

Risk factors for osteoporosis are many, and include :

- genetic predisposition
- poor nutrition in childhood and adolescence
- early age of onset of menopause
- decreased exposure to sunshine and low vitamin D intake
- low dietary calcium
- alcohol and cigarette use
- caffeine consumption
- low weight
- lack of regular exercise

References

1 MC Chapuy, ME Arlot, F Debouef etc. Vitamin D3 and calcium to prevent hip fractures in elderly women. New England Journal of Medicine 1992 327: 1637-42.

2 MC Chapuy, ME Arlot, PD Delmas, PJ Meunier. Effect of calcium and cholecalciferol treatment for three years on hip fractures in elderly women.

3 WJ Gillespie, DA Henry, DL O'Connell, J Robertson. Vitamin D and Vitamin D analogues in the prevention of fractures in involutional and post-menopausal osteoporosis. The Cochrane Library 1997, issue 1.

4 ***Bandolier*** 25, March 1996.

Analysis	NNT at 18 months	NNT at 36 months
Complete follow up		
All fractures	33 (16 - 142)	
Hip fractures	50 (29 - 881)	
Active-treatment analysis		
All fractures	20 (13 - 48)	14 (9 - 38)
Hip fractures	33 (20 - 150)	20 (12 - 65)
Intention to treat analysis		
All fractures	25 (16 - 79)	14 (9 - 30)
Hip fractures	50 (25 - 485)	20 (13 - 57)

MINDSTRETCHER

Logistic regression explained

A tutorial from Professor Peter Freeman in sunny Cornwall:

So ***Bandolier*** confesses to finding logistic regression analysis impenetrable. Well said! So, I suspect, does the majority of the medical community, so let's have a go at explaining.

The example that ***Bandolier*** reported was typical. There was one outcome variable, liver cirrhosis confirmed by biopsy, and a whole host of explanatory variables, clinical observations as well as biochemistry results. The aim was to find a subset of all the explanatory variables that can be combined to predict the value of the outcome variable. This is the usual technique that statisticians call regression analysis. It is done using tons of complicated arithmetic, but the outcome is a regression equation having the outcome variable on the left hand side and a combination of the explanatory variables on the right. For any future patient, the values of their explanatory variables can be fed into the equation to predict the value of their outcome variable.

The important result was that liver cirrhosis can be predicted with a high accuracy using selected clinical observations only and that biochemistry readings are of little or no extra help in the diagnosis. Even better, only six of the 25 clinical factors are important, so good diagnosis only needs observation of abdominal wall veins, facial telangiectasia, fatness, peripheral oedema, vascular spiders and white nails.

For technically-minded readers who are curious about the mysterious word "logistic" that pops up before "regression", this is simply a little mathematical trick to get over the snag that in this example the risk of liver cirrhosis can only take values between 0 and 1. Something called a logistic transformation is used to change this into something more numerically convenient, with a reverse transformation to get us back again to the real world afterwards.

Some sums (not too hard)

Readers who are still game for more enlightenment might like to try working through the following details:-

Table 3 of the paper Hamberg *et al* [1] gives the clinical model logistic regression equation as

LOGODDS = -4.18 + 3.01xFAC + 1.80xVAS + 1.75xWHI + 1.48xABD + 1.07xFAT - 1.28xPER

where:

FAC	=	facial telangiectasia
VAS	=	vascular spiders
WHI	=	white nails
ABD	=	abdominal wall veins
FAT	=	fatness
PER	=	peripheral oedema

and the funny word on the left-hand side of the equation just stands for the mathematical way of writing the logistic transformation.

Now I think (although the paper does not say so explicitly) that the variables on the right-hand side take the value 1 if the clinical sign is present and the value 0 if it is absent. If this is so, then a patient who showed all six clinical signs would have a risk of liver cirrhosis given by

LOGODDS = -4.18 + 3.01 + 1.80 +1.75 + 1.48 + 1.07 - 1.28

= 3.65

and so

ODDS = $e^{3.65}$ = 38.47
(any scientific calculator has a button for this)

and hence

RISK = ODDS / (1 + ODDS) = 38.47/39.47 = 0.97 or 97%

while a patient who showed only white nails and fatness would have risk

LOGODDS = -4.18 + 1.75 + 1.07 = -1.36

ODDS = $e^{-1.36}$ = 0.26

RISK = 0.26 / 1.26 = 0.20 or 20%

QED

Professor Peter Freeman
Ship Cottage, Cadgwith
Helston, Cornwall TR12 7JX

Reference:
1 KJ Hamberg, B Carstensen, T Sørensen, K Eghøje. Accuracy of clinical diagnosis of cirrhosis among alcohol-abusing men. Journal of Clinical Epidemiology 1996 49: 1295-1301.

Bandolier enjoyed that and now feels much more confident about using such logistic regression information. It looks straightforward enough to program onto something like Excel or on a personal organiser.

SCREENING FOR PROSTATE CANCER

Some of the jewels in the NHS R&D Directorate crown were apparent very early in this great experiment - the Cochrane Collaboration, for example, with the Cochrane Library, and the Centre for Reviews and Dissemination at York with Effectiveness Bulletins and Effectiveness Matters. The Health Technology Assessment programme has taken a little longer to mature because large scale reviews and trials from the process begun in 1994 are only now available.

The first three reviews (two on prostate cancer) have just been published [1-3]. ***Bandolier*** is impressed with the results of the two it has read, on prostate cancer. The reviews, while covering similar territory, were done by different teams. Importantly, they reach the same conclusion, that screening for prostate cancer is unproven, and is unlikely to be of any benefit.

In the first ***Bandolier*** conference on mens' health, the enthusiastic US attitude about screening was laid out as the key problem in prostate cancer being to screen early enough to find the younger patients who need treatment without frightening lots of normal people. Screening should not be undertaken in old men. A good rule was not to do PSAs on men over 70 "unless they bring both parents with them".

That's fine, but it assumes effectiveness of both diagnosis and treatment. The reviews show just how uncertain are the diagnostic methods (rectal examination and PSA) we have. The active treatment options of radical prostatectomy and radiotherapy both have a significant mortality, and huge morbidity in both urinary incontinence and impotence. For men in their '50s this constitutes a big downside, well laid out in the Effectiveness Matters document from York for men asking for PSA tests (perhaps after reading it some of our enthusiastic journalist docs in the national newspapers may become less enthusiastic). It's an even bigger downside when there seems to be no evidence that the cost can be set against more years of life.

The cost of a full-scale screening programme for the UK is estimated at between £500 million to £1,500 million. It would be money down the drain.

Babies and bathwater

Bandolier has heard it suggested that some Trusts will stop all PSA testing on seeing these reports. That would be a mistake. There are good reasons for using PSA in prostate cancer treatment. It can help make judgements about prognosis and treatment [4].

Getting a copy

Most GPs will have had an information pack with Effectiveness Matters and the superlative companion Effectiveness Matters for patients who might ask for a PSA test. The two monographs are available from the National Co-ordinating Centre for HTA (Fax +44 (0) 1962 877425 at the high cost of £50 each. ***Bandolier*** believes these should be on most GPs' shelves, and if enough of you were to express an interest so that a higher print run could be contemplated, the cost would surely drop to something much more affordable.

References:
1 DM Richards, JJ Deeks, TA Sheldon, JL Shaffer. Home parenteral nutrition. Health Technology Assessment 1997 1: no1.
2 S Selley, J Donovan, A Faulkner, J Coast, D Gillatt. Diagnosis, management and screening of early localised prostate cancer. Health Technology Assessment 1997 1: no2.
3 J Chamberlain, J Melia, S Moss, J Brown. The diagnosis, management, treatment and costs of prostate cancer in England and Wales. Health Technology Assessment 1997 1: no3.
4 RA Moore. Prostate-Specific Antigen. Kaleidescope 4: 1-55 (1994).

EDITORS

Dr Andrew Moore
Dr Henry McQuay Dr J A Muir Gray

Pain Relief Unit
The Churchill, Oxford OX3 7LJ

Editorial office: 01865 226132
Editorial fax: 01865 226978
Email: andrew.moore@pru.ox.ac.uk
Internet: http://www.jr2.ox.ac.uk/Bandolier

ISSN 1353-9906

Fatality rates in Britain (mainly from 1992).

Deaths per 100 million passenger:

Mode of transport	journeys	hours	km
Motorbike	100.0	300.0	9.70
Air	55.0	15.0	0.03
Water	25.0	12.0	0.60
Bicycle	12.0	60.0	4.30
Foot	5.1	20.0	5.30
Car	4.5	15.0	0.40
Van	2.7	6.6	0.20
Rail	2.7	4.8	0.10
Bus/Coach	0.3	0.1	0.04

Data from Royal Society of Prevention of Accidents